$$\frac{1}{1+e^{-z}} = f(z)$$

$$f'(z) = \frac{-1}{(1+e^{-z})^2} \cdot e^{-z} \cdot (-1)$$

$$= f(z)$$

$$\boxed{= \frac{e^{-z}}{(1+e^{-z})^2}} = \left(\frac{1}{1+e^{-z}}\right)\left(1 - \frac{1}{1+e^{-z}}\right)$$

$$= f(z)\,(1 - f(z))$$

RMS prop

$$W_{n+1} = W_n - \frac{\boxed{\gamma}\,\nabla_w L}{|C_{n+1}|}$$

$\boxed{\text{adam}}$

$$\boxed{C_{n+1} = \beta(C_n) + (1-\beta)\,\nabla_w L}$$

momentum.

$$W_{n+1} = W_n - \gamma V_{n+1}$$

$$\boxed{V_{n+1}} = \rho V_n + \nabla_w L \qquad \rho = 0.9$$

$$y_1 = \underbrace{W_1}\,^{T} \underbrace{X} + \underbrace{b_1}$$

$$\underset{n}{y_2} = W_2^{T} \underbrace{y_1}_{m} + b_2$$

$$\cdots$$

$$y_n = W_n^{T} y_{n-1} + b_{n-1}$$

$$y = w^T x + b = \sum_{i=0}^{n} w_i x_i + b$$

$$Var(y) - ?$$

$$Var(y) = \mathbb{E}(w x)^2 - (\mathbb{E} w x)^2$$

$$= \mathbb{E}(w)^2 \mathbb{E} x^2 - (\mathbb{E} w \mathbb{E} x)^2$$

$$= (\mathbb{E} w^2 - (\mathbb{E} w)^2 + (\mathbb{E} w)^2)(\mathbb{E} x^2 - (\mathbb{E} x)^2 + (\mathbb{E} x)^2) - (\mathbb{E} w \mathbb{E} x)^2 =$$

$$= (Var\, w + (\mathbb{E} w)^2)(Var\, x + (\mathbb{E} x)^2) - (\mathbb{E} w \mathbb{E} x)^2$$

$$= \operatorname{Var} w \operatorname{Var} x + \cancel{\operatorname{Var} x (\mathbb{E} w)^2} +$$

$$+ \cancel{\operatorname{Var} w (\mathbb{E} x)^2} + \cancel{\mathbb{E} x^2 \mathbb{E} w^2} -$$

$$- \cancel{\mathbb{E} x^2 \mathbb{E} w^2}$$

$$\boxed{w \sim \mathbb{N}(0, 1).}$$

$$\operatorname{Var}(y_i) = \underline{\operatorname{Var}(w) \cdot \operatorname{Var}(x).}$$

$$\operatorname{Var}(y) = \sum_{i=0}^{k} \operatorname{Var}(y_i) = k \operatorname{Var}(y_i) =$$

$$= \boxed{k \operatorname{Var}(w)} \, \underline{\operatorname{Var}(x).}$$

$$\boxed{w \sim U\left[-\frac{1}{\sqrt{k}} ; +\frac{1}{\sqrt{k}}\right]}$$

$$\operatorname{Var}(w) = \frac{1}{12} \left(\frac{2}{\sqrt{k}} \right)^2 = \frac{1}{3k}$$

$$Var(y) = k \cdot \frac{1}{3 \cdot k} \cdot Var(x) =$$

$$= \frac{1}{3} Var(x).$$

$$Var(\hat{y}) = \left(\frac{1}{3}\right)^{100} \cdot Var(x) \approx 0$$

$$Var(w) = \frac{2}{k + m}$$

$$W \sim U\left[-\frac{\sqrt{6}}{\sqrt{k+m}}, \frac{\sqrt{6}}{\sqrt{k+m}}\right]$$

$$\boxed{\text{Glorot Uniform}}$$

$\boxed{\text{He.}}$ $W \sim N\left(0, \sqrt{\frac{2}{n}}\right)$

$W \sim U(?, ?)$ $\sqrt{n}$ $\boxed{k}$

$Var(y) = n\,Var(w)\,Var(x)$

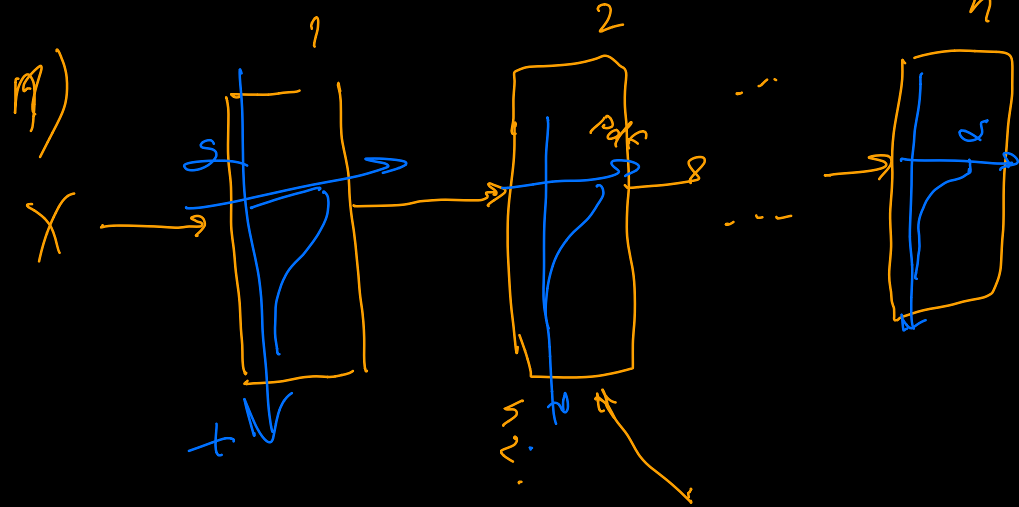
1) activation. ; Relu.

2) Xavier, He.

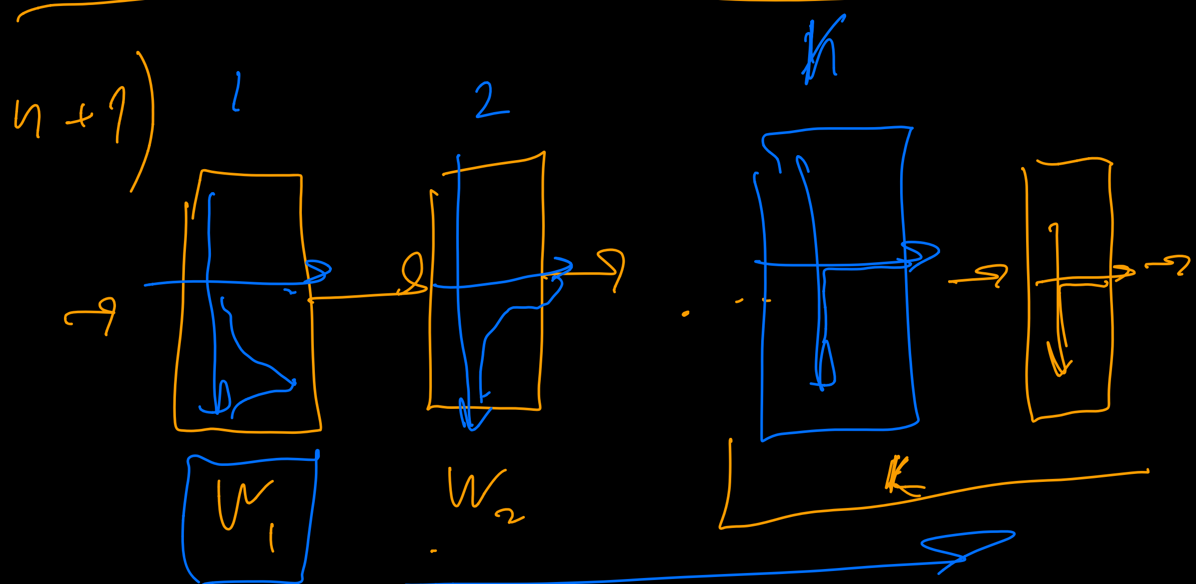
3) $\boxed{\text{nan.}}$; gradient clip. tf.clip.

4) $lr_n = lr_{n-1} \cdot \alpha$, $\alpha < 1$.

5).



$\boxed{h_1} = f(W_1^T \cdot x)$ $\quad \boxed{h_2} = f(W_2^T \cdot h_1)$



$$\hat{h}_1 = \frac{h_1 - \mathbb{E}h_1}{std(h_1) + \varepsilon}$$

$$\boxed{\mu_n} = \mu_{n-1} \cdot p + (1-p) \cdot \mathbb{E}h_1$$

$$\frac{1}{m} \sum_{i=1}^{\boxed{m}} h_1^{(i)} = \mathbb{E}h_1$$

$$\boxed{\xi_n} = \xi_{n-1} \cdot p + (1-p) \cdot std(h_1)$$

$$\tanh\left(bn\left(W^T h_{k-1}\right)\right)$$

$$\hat{h}_1 = \boxed{\frac{h_1 - \mathbb{E}h_1}{std(h_1)+\varepsilon} \cdot \boxed{z} + \boxed{y}} \approx 0$$